

248

**Reihe Ökonomie
Economics Series**

Cointegration Analysis with State Space Models

Martin Wagner

248

**Reihe Ökonomie
Economics Series**

Cointegration Analysis with State Space Models

Martin Wagner

February 2010

**Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna**

Contact:

Martin Wagner
Institute for Advanced Studies
Department of Economics and Finance
Stumpergasse 56
A-1060 Vienna, Austria
☎: +43/1/599 91-150
Fax: +43/1/599 91-555
email: Martin.Wagner@ihs.ac.at
and
Frisch Centre for Economic Research
Oslo, Norway

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share “work in progress” in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschaftswissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

Abstract

This paper presents and exemplifies results developed for cointegration analysis with state space models by Bauer and Wagner in a series of papers. Unit root processes, cointegration and polynomial cointegration are defined. Based upon these definitions the major part of the paper discusses how state space models, which are equivalent to VARMA models, can be fruitfully employed for cointegration analysis. By means of detailing the cases most relevant for empirical applications, the $I(1)$, $MFI(1)$ and $I(2)$ cases, a canonical representation is developed and thereafter some available statistical results are briefly mentioned.

Keywords

State space models, unit roots, cointegration, polynomial cointegration, pseudo maximum likelihood estimation, subspace algorithms

JEL Classification

C13, C32

Contents

1	Introduction	1
2	Unit Roots and Cointegration	2
3	Cointegration Analysis with State Space Models	6
	3.1 Structure Theory	7
	3.2 Statistical Theory	19
4	Open Issues, Summary and Conclusions	24
	Acknowledgements	25
	References	25

1 Introduction

Since the seminal contribution of Clive W.J. Granger (1981) that introduced the concept of cointegration, the modelling of multivariate (economic) time series using methods and models that allow for and incorporate unit roots and cointegration has become standard econometric practice with applications in many areas ranging from macroeconomics to finance.¹ Cointegration refers to the fact that linear combinations of unit root – also referred to as integrated – processes can be integrated of a lower order than the processes themselves (precise definitions of unit root processes and cointegration are given in Section 2).

Cointegration analysis is performed using several approaches that can be roughly divided in non-parametric and parametric modelling. The non-parametric approach dates back to Engle and Granger (1987) and in this approach the focus is only on testing for and estimating the cointegrating relationships whilst all other characteristics of the data generating process (DGP) are treated as nuisance parameters. Contributions pursuing this approach include (from a long list) Bierens (1997), Park and Phillips (1988, 1989), Phillips and Hansen (1990), Saikkonen (1991), Sims et al. (1990) or Stock and Watson (1988). As the name suggests, in parametric analysis a fully specified model class is posited and cointegration analysis is performed within this framework. The by far most prominent parametric model class for cointegration analysis are vector autoregressive (VAR) models, popular due to the important work of Søren Johansen and his co-authors (see his monograph Johansen 1995).²

Some authors have also considered vector autoregressive moving average (VARMA) models for cointegration analysis, e.g. Yap and Reinsel (1995) and Lütkepohl and Claessen (1997). This allows (in principle) to overcome some potentially relevant limitations of VAR models including: First, it is well-known since Zellner and Palm (1974) that processes composed of subsets of the variables of VAR processes in general follow VARMA processes (with the empirical literature full of examples where also for subsets of variables for which VAR models have been fitted VAR models are considered).³ Second, quite similarly also aggregation of VAR processes leads to VARMA

¹Clive W.J. Granger and Robert F. Engle shared the Nobel prize in economics in 2003. One of the contributions for which they have been awarded is cointegration. The second awarded contribution are so-called ARCH models that allow to model time-varying conditional variances, a pertinent phenomenon in e.g. financial time series. Note as a historical remark that several other researchers also were ‘close to discovering’ cointegration around the same time, e.g. Box and Tiao (1977) or Krämer (1981).

²His work on cointegration analysis with VAR models has made Søren Johansen one of the most widely cited econometricians. VAR models are not only the by far most popular approach in cointegration analysis, VAR cointegration analysis is also the ‘most developed’ strand of the literature, providing a large battery of tools for diagnostic testing (including stability testing) as well as other tools that are considered useful for empirical analysis, e.g. impulse response analysis and forecast error variance decompositions.

³In order to address this problem the literature often refers to approximation results that show that VARMA, or in fact even more general, processes can be approximated by VAR processes with the lag lengths tending to infinity with the sample size at certain rates. This line of work goes back to Lewis and Reinsel (1985) for stationary processes. Extensions of some form or another to cointegrated processes are provided by Saikkonen (1992), Saikkonen and

processes. Temporal aggregation is e.g. a relevant concern when studying the term structure of interest rates with cointegration methods, for a detailed discussion see Bauer and Wagner (2009). Third, the (linearized) solutions to dynamic stochastic general equilibrium models are typically VARMA rather than VAR processes, as has been pointed out already in Campbell (1994). Fourth, a VARMA model may be a more parsimonious description of the DGP than a VAR model.

State space models (as considered here) are an equivalent model class to VARMA models (a detailed discussion of state space models and their links to VARMA models in the stationary case is contained in Hannan and Deistler 1988).⁴ Like VARMA models state space models have rarely been used for cointegration analysis. A few early exceptions include Aoki (1987) and Aoki and Havenner (1989, 1991), all of which deal with the $I(1)$ case only. In a series of papers Bauer and Wagner develop structure theory as well as statistical theory for cointegration analysis with state space models. Some of their contributions are discussed in the course of this paper.

The paper is organized as follows: In Section 2 unit root processes, cointegration and polynomial cointegration are defined. Section 3 is devoted to a discussion of cointegration analysis with state space models, where both structure theoretic as well as statistical aspects are considered. Finally, Section 4 very briefly mentions some open questions that need to be addressed to render cointegration analysis with state space models a fully fledged alternative to VAR cointegration analysis and this section also concludes. This paper does not offer new theoretical results, instead it is intended to give an overview of cointegration analysis with state space models at one place. Correspondingly, the presentation at times will be exemplary and a bit sloppy to convey the main ideas rather than to present results in their most general or exact form. Throughout, however, references to precise discussions will be provided.

2 Unit Roots and Cointegration

Since the late 1970s the question of (trend-) stationarity versus unit root nonstationarity of economic and financial time series has received a lot of attention. E.g. in macroeconomics an important contribution (that spurred a lot of interest of both economists and econometricians) has been the work of Nelson and Plosser (1982), who study the trend behavior of 14 US macroeconomic time series. They find that all but one of the series have a stochastic rather than a deterministic trend, i.e. are unit root processes, the exception being the unemployment rate (for which to test

Luukkonen (1997) and Bauer and Wagner (2007). On top of the issue of the existence and properties of such a sequence of VAR approximations (as a function of sample size) also the question whether a VAR approximation is parsimonious arises.

⁴Here we focus only on linear state space models and use a similar approach as Hannan and Deistler (1988) or Kailath (1980). In the econometric literature the term state space model is also used differently and sometimes more generally, see e.g. Durbin and Koopman (2001), Harvey (1989) or Kim and Nelson (1999).

for a unit root may be unnecessary on conceptual grounds in the first place).⁵ The nature of the trend component in time series is not only of interest (to stick to the example) for macroeconomists trying to build models with corresponding propagation mechanisms that reflect the trend components of observed series (or instead use correspondingly filtered time series), but also has important consequences for statistical analysis. It is known since a long time (see for example the presidential address to the Royal Statistical Society of Yule 1926) that the presence of unit roots or stochastic trends has drastic consequences for the behavior of regressions. A regression of two stochastically independent random walks on each other leads to a seemingly significant regression coefficient, a phenomenon labeled nonsense-regression by Yule (1926) and spurious regression by Granger and Newbold (1974). The latter paper provides simulation evidence only and an analytical study of spurious regression and its asymptotic properties is given in Phillips (1986). It turns out, for the mentioned regression of two independent random walks on each other, that the limit of the regression coefficient is non-zero and that its ('textbook OLS') t -value diverges with rate square root of sample size, thus resulting in seemingly significant coefficients also and especially in large samples. The understanding of spurious regression led to concerns about the validity of findings involving potentially unit root nonstationary variables obtained with 'traditional' methods, which in turn ignited the large and ongoing research in cointegration analysis which allows for valid inference in unit root nonstationary settings. In particular, as we shall see below, cointegration is equivalent to common stochastic trends and cycles being present in some (or all) of the variables and thus is a convenient way of describing the relationships between unit root nonstationary variables. In particular testing for the presence (respectively) absence of cointegration allows to distinguish a spurious regression situation from a situation in which the variables studied are indeed related.

Let us now turn to formal definitions of unit roots and cointegration, following Bauer and Wagner (2005). We consider s -dimensional (real valued) stochastic processes $(y_t)_{t \in \mathbb{Z}}$ with zero mean, i.e. $\mathbb{E}(y_t) = 0$, $t \in \mathbb{Z}$. The difference operator at frequency $0 \leq \omega \leq \pi$ is defined as

$$\Delta_\omega(L) := \begin{cases} 1 - e^{i\omega}L, & \omega \in \{0, \pi\} \\ (1 - e^{i\omega}L)(1 - e^{-i\omega}L), & \omega \in (0, \pi). \end{cases} \quad (1)$$

Here L denotes the backward shift operator, such that $L(y_t)_{t \in \mathbb{Z}} = (y_{t-1})_{t \in \mathbb{Z}}$. Keeping this definition in mind we also use the sloppy shorthand notation $Ly_t = y_{t-1}$, with which we obtain e.g. $\Delta_\omega(L)y_t = y_t - 2 \cos(\omega)y_{t-1} + y_{t-2}$ for $0 < \omega < \pi$. Note that for $0 < \omega < \pi$ the difference operator $\Delta_\omega(L)$ filters the pair of complex conjugate unit roots $e^{\pm i\omega}$ jointly. This ensures that also filtered processes $\Delta_\omega(L)(y_t)_{t \in \mathbb{Z}}$ are real valued for real valued processes $(y_t)_{t \in \mathbb{Z}}$.

⁵Nelson and Plosser (1982) employed the unit root tests of Dickey and Fuller (1981), which are amongst the most widely-used unit root tests. The unit root and stationarity testing literature has since then grown to an extremely large literature with numerous tests developed under various sets of assumptions against all sorts of alternatives (e.g., nonlinear alternatives, alternatives with deterministic trend components with breaks, etc.).

Definition 1 The s -dimensional zero mean process $(y_t)_{t \in \mathbb{Z}}$ has unit root structure

$$\Omega := ((\omega_1, h_1), \dots, (\omega_l, h_l)),$$

with $0 \leq \omega_1 < \omega_2 < \dots < \omega_l \leq \pi, h_k \in \mathbb{N}, k = 1, \dots, l$, if

$$D(L)y_t := \Delta_{\omega_1}^{h_1}(L) \cdots \Delta_{\omega_l}^{h_l}(L)y_t = v_t, \quad t \in \mathbb{Z}, \quad (2)$$

with $v_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ a linearly regular⁶ stationary process. Here $c(z) := \sum_{j=0}^{\infty} c_j z^j, z \in \mathbb{C}$ with $c_j \in \mathbb{R}^{s \times s}, j \geq 0$ and $\sum_{j=0}^{\infty} \|c_j\| < \infty$ corresponds to the Wold representation of $(v_t)_{t \in \mathbb{Z}}$. It holds that $c(e^{i\omega_k}) \neq 0$ for $k = 1, \dots, l$ and $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a white noise process with $\mathbb{E}\varepsilon_t \varepsilon_t' = \Sigma > 0$.

Processes $(y_t)_{t \in \mathbb{Z}}$ with nonempty unit root structure are called unit root processes. The set $\{\omega_1, \dots, \omega_l\}$ is referred to as set of unit root frequencies and the integers $h_k, k = 1, \dots, l$ are called integration orders. A (zero mean) linearly regular stationary process is said to have empty unit root structure $\Omega_0 := \{\}$.

A unit root process with unit root structure $((0, m))$, $m \in \mathbb{N}$, is called $I(m)$ process and a unit root process with unit root structure $((\omega_1, 1), \dots, (\omega_l, 1))$ is called multiple frequency $I(1)$, in short MFI(1), process.

As discussed in Bauer and Wagner (2005, Section 2) the unit root structure as just defined is unique. Note furthermore that in the literature definitions of integrated processes $(z_t)_{t \in \mathbb{Z}}$ are often formulated for demeaned processes, i.e. for processes $(z_t - \mathbb{E}(z_t))_{t \in \mathbb{Z}}$, compare e.g. Johansen (1995, Definition 3.3, p. 35). Here we only consider the latter zero mean process. For further discussion concerning the definition of unit root processes see Bauer and Wagner (2005, Section 2).

Before we now turn to a definition of cointegration and polynomial cointegration we first define a semi-ordering of unit root structures. Denote for a unit root structure Ω with $F(\Omega)$ the set of distinct unit root frequencies included, i.e. $F(\Omega) := \{\omega_1, \dots, \omega_l\}$.

Definition 2 Let $\tilde{\Omega} := ((\tilde{\omega}_1, \tilde{h}_1), \dots, (\tilde{\omega}_{\tilde{l}}, \tilde{h}_{\tilde{l}}))$ and $\Omega := ((\omega_1, h_1), \dots, (\omega_l, h_l))$ be two unit root structures. Then $\tilde{\Omega} \preceq \Omega$ if

1. $F(\tilde{\Omega}) \subseteq F(\Omega)$.
2. For all $\omega \in F(\tilde{\Omega})$ for \tilde{k} and k such that $\tilde{\omega}_{\tilde{k}} = \omega_k = \omega$ it holds that $\tilde{h}_{\tilde{k}} \leq h_k$.

Further $\tilde{\Omega} \prec \Omega$ if $\tilde{\Omega} \preceq \Omega$ and $\tilde{\Omega} \neq \Omega$.

For two unit root structures $\tilde{\Omega} \preceq \Omega$ define the decrease $\delta_k(\Omega, \tilde{\Omega})$ of the integration order at frequency

⁶For a definition see Hannan and Deistler (1988, p. 20).

ω_k , for $k = 1, \dots, l$, as

$$\delta_k(\Omega, \tilde{\Omega}) := \begin{cases} h_k - \tilde{h}_{\tilde{k}} & , \quad \exists \tilde{k} : \tilde{\omega}_{\tilde{k}} = \omega_k \in F(\tilde{\Omega}), \\ h_k & , \quad \omega_k \notin F(\tilde{\Omega}). \end{cases}$$

Clearly, the empty unit root structure is the smallest element in this semi-ordering, i.e. $\Omega_0 := \{\} \prec \Omega$ for any non-empty unit root structure Ω .

The definitions of cointegration and polynomial cointegration are based on the introduced semi-ordering of unit root structures.

Definition 3 *An s -dimensional unit root process $(y_t)_{t \in \mathbb{Z}}$ with unit root structure $\Omega = ((\omega_1, h_1), \dots, (\omega_l, h_l))$ is called cointegrated of order $(\Omega, \tilde{\Omega})$, where $\tilde{\Omega} \prec \Omega$, if there exists a vector $\beta \in \mathbb{R}^s, \beta \neq 0$, such that $(\beta' y_t)_{t \in \mathbb{Z}}$ has unit root structure $\tilde{\Omega}$. The vector β is in this case called cointegrating vector (CIV) of order $(\Omega, \tilde{\Omega})$.*

Until now by far most applications of cointegration analysis are concerned with the analysis of I(1) processes. Closely related is, by considering unit roots with integration orders equal to 1 also at non-zero frequencies, the case of so-called seasonal unit roots and seasonal cointegration has also received quite some attention. In our notation this corresponds to the case of MFI(1) processes with the unit root frequencies corresponding to the ‘seasonal’ frequencies. In the case of quarterly data thus the considered unit root frequencies are $0, \pi/2$ and π . Early contributions in seasonal cointegration analysis include Hylleberg, Engle, Granger and Yoo (1990), Lee (1992) and Ahn and Reinsel (1994).

The concept of cointegration has been generalized by considering not only static but also dynamic linear transformations of the form $\sum_{j=0}^q \beta'_j y_{t-j}$. Such a generalization has first been introduced as multi-cointegration by Yoo (1986) and Granger and Lee (1989a, 1989b). Multi-cointegration prevails for an s -dimensional I(2) process $(y_t)_{t \in \mathbb{Z}}$, if there exists a linear combination of the two I(1) processes $(\beta' y_t)_{t \in \mathbb{Z}}$ – i.e. $\beta \in \mathbb{R}^{s \times k}, k \geq 1$, comprises one or more cointegrating vectors that reduce the integration order from 2 to 1 – and $(y_t - y_{t-1})_{t \in \mathbb{Z}}$ that is stationary. The concept of multi-cointegration has been generalized (by allowing for higher integration orders and higher order linear dynamic transformations) to polynomial cointegration by Gregoir and Laroque (1994) with one formal definition given in Gregoir (1999). We shall see below that already in the MFI(1) case a certain form of polynomial cointegration, referred to as dynamic cointegration, may occur that annihilates the stochastic cycles to unit root frequencies $0 < \omega < \pi$.

As discussed in Bauer and Wagner (2005, Section 5) and in more detail in Bauer and Wagner (2008) the definition of Gregoir (1999, Definition 3.1) has several limitations and drawbacks, including that it does not generalize multi-cointegration. Our definition of polynomial cointegration given

next has as one of its advantages that it does generalize multi-cointegration (for further discussion see Bauer and Wagner 2008).

Definition 4 *An s -dimensional unit root process $(y_t)_{t \in \mathbb{Z}}$ with unit root structure Ω is called polynomially cointegrated of order $(\Omega, \tilde{\Omega})$, where $\tilde{\Omega} \prec \Omega$, if there exists a vector polynomial $\beta(z) = \sum_{m=0}^q \beta_m z^m$, $\beta_m \in \mathbb{R}^s$, $m = 0, \dots, q$, $\beta_q \neq 0$ for some integer $1 \leq q < \infty$ such that:*

(i) $\beta(L)'(y_t)_{t \in \mathbb{Z}}$ has unit root structure $\tilde{\Omega}$,

(ii) $\max_{k=1, \dots, l} \|\beta(e^{i\omega_k})\| \delta_k(\Omega, \tilde{\Omega}) \neq 0$.

The vector polynomial $\beta(z)$ is in this case called polynomial cointegrating vector (PCIV) of order $(\Omega, \tilde{\Omega})$.

The restriction formulated in item (ii) of the above definition excludes vector polynomials that reduce the integration orders by merely differencing the process.

It is clear that a unit root process $(y_t)_{t \in \mathbb{Z}}$ can be cointegrated and polynomially cointegrated of different orders. However, for any CIV or PCIV for a process $(y_t)_{t \in \mathbb{Z}}$, its cointegration respectively polynomial cointegration order is unique. This follows from the above mentioned uniqueness of the unit root structure of unit root processes. Note furthermore that not every statically or dynamically linearly transformed unit root process needs to be a unit root process. Components in the kernel of the differencing filter $D(L)$ as given in Definition 1 may cause problems, for details see Bauer and Wagner (2005).

As concerns applications, the seasonal MFI(1) and I(2) cases are the prime cases in which polynomial cointegration has been studied for economic phenomena. I(2) cointegration analysis has been developed and applied e.g. in Johansen (1997) or Stock and Watson (1993).

3 Cointegration Analysis with State Space Models

A stochastic process $(y_t)_{t \in \mathbb{Z}}$, $y_t \in \mathbb{R}^s$ is said to have a state space representation if it is a solution to the *state space* or *system* equations

$$\begin{aligned} y_t &= Cx_t + \varepsilon_t, \\ x_{t+1} &= Ax_t + B\varepsilon_t, \end{aligned} \tag{3}$$

for a white noise process $(\varepsilon_t)_{t \in \mathbb{Z}}$. The unobserved process $(x_t)_{t \in \mathbb{Z}} \in \mathbb{R}^n$ is referred to as state process and $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times s}$ and $C \in \mathbb{R}^{s \times n}$ real matrices. The triple (A, B, C) is referred to as a state space realization of the *transfer function* $k(z) := I_s + zC(I_n - zA)^{-1} = I_s + \sum_{j=1}^{\infty} CA^j B z^j$, well defined for $z \in \mathbb{C}$ such that $|z| < (|\lambda_{\max}(A)|)^{-1}$, where $\lambda_{\max}(A)$ denotes an eigenvalue of

maximum modulus of the matrix A . The transfer function characterizes the dependence of $(y_t)_{t \in \mathbb{Z}}$ on $(\varepsilon_t)_{t \in \mathbb{Z}}$ and thus summarizes the dynamic properties of $(y_t)_{t \in \mathbb{Z}}$.

The above format of a state space system with the same noise process $(\varepsilon_t)_{t \in \mathbb{Z}}$ appearing in both the output and state equation is often referred to as innovations form. It can be shown that any state space model of the form $y_t = Cx_t + \varepsilon_t$, $x_{t+1} = Ax_t + v_t$ can be transformed into innovation form, see e.g. Aoki and Havenner (1991, Section 2).

The sequence of the coefficients of the transfer function is, especially in the VAR literature, referred to as impulse response sequence. Thus, computation of the impulse response sequence is a trivial task for state space models once the system matrices (A, B, C) have been estimated, with the impulse response coefficient matrices given by I_s, CB, CAB, \dots

As discussed (for the stationary case) in great detail in Hannan and Deistler (1988, Chapters 1 and 2) there are close connections between state space models and VARMA models, which are essentially equivalent model classes. Lemma 1 in Bauer and Wagner (2005) discusses the equivalence of state space and VARMA models also in the unit root case. Thus, both model classes allow to study *rational* unit root processes. Using the notation of Definition 1 by this we mean that the transfer function $c(z)$ corresponding to the process $(v_t)_{t \in \mathbb{Z}}$ is restricted to be a rational function, i.e. $c(z) = a^{-1}(z)b(z)$, with $a(z)$ and $b(z)$ left co-prime matrix polynomials. The pair of matrix polynomials $(a(z), b(z))$ is referred to as VARMA realization of the transfer function $c(z)$, for which as just mentioned also state space realizations (A, B, C) exist (see also the discussion in Hannan and Deistler 1988, Chapters 1 and 2).

Remark 1 *The discussion in Bauer and Wagner (2006) indicates, for the MFI(1) case, how state space modeling ideas can be used to approximate non-rational unit root processes. For rational approximation of stationary processes see also Hannan and Deistler (1988, Chapter 7.4). These aspects are not pursued further in this paper where we only consider rational unit root processes.*

3.1 Structure Theory

Like VARMA models also state space models for a given rational process $(y_t)_{t \in \mathbb{Z}}$ are not unique, i.e. there exist multiple state space realizations $(A_i, B_i, C_i)_{i \in \mathcal{I}}$ that correspond to the same transfer function. Denote with S_n the set of all triples of system matrices (A, B, C) with $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times s}$ and $C \in \mathbb{R}^{n \times n}$ and denote with U the set of all $s \times s$ rational transfer functions $k(z)$ with $k(0) = I_s$. Now define a mapping $\Pi : \bigcup_{n \geq 0} S_n \rightarrow U$ attaching the transfer function $k(z)$ to the triple $(A, B, C) \in S_n$ (as given below (3) above) for $n \geq 0$. Using this notation, non-uniqueness of state space realizations refers to the fact that the pre-image of the mapping Π for a

given transfer function $k(z)$ is generally not a singleton. *Identification* of state space systems thus requires to impose a sufficient set of restrictions on (A, B, C) such the pre-image of the mapping Π only contains one element for each $k(z) \in \tilde{U}$, where \tilde{U} denotes the set of transfer functions for which identification is considered.

There are two sources of non-uniqueness of state space realizations for a given transfer function. First, there exist state space realizations with different state dimensions. A state space system $(A, B, C) \in S_{\tilde{n}}$ is called *minimal*, if there exists no state space system $(A, B, C) \in S_{\tilde{n}}$ such that $\Pi(A, B, C) = \Pi(\tilde{A}, \tilde{B}, \tilde{C})$ and $n > \tilde{n}$, i.e. if the state dimension is minimal. Minimality is the state space analogue to left coprimeness in the VARMA framework. The dimension n of the state in a minimal state space model is called order of the state space model or order of the transfer function. Second, non-uniqueness arises via the choice of the basis of the state, for fixed state dimension. It holds that two minimal state space realizations $(A, B, C) \in S_n$ and $(\tilde{A}, \tilde{B}, \tilde{C}) \in S_n$ are *observationally equivalent*, i.e. $\Pi(A, B, C) = \Pi(\tilde{A}, \tilde{B}, \tilde{C})$, if and only if there exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that $\tilde{A} = TAT^{-1}$, $\tilde{B} = TB$ and $\tilde{C} = CT^{-1}$. Note that such a transformation implies a corresponding basis change of the state vector to $\tilde{x}_t = Tx_t$.

Minimality captures the fact that there are no ‘superfluous’ components contained in the state vector that do not influence the output. This in turn implies that minimality is a necessary condition for correspondence of the unit root properties of the state process and the output process (for details see Bauer and Wagner 2005, Theorem 3). The main idea is, however, also immediately seen in a simple example of a non-minimal system:

$$\begin{aligned} y_t &= \begin{bmatrix} C_1 & 0 \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,2} \end{bmatrix} + \varepsilon_t \\ \begin{bmatrix} x_{t+1,1} \\ x_{t+1,2} \end{bmatrix} &= \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,2} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \varepsilon_t \end{aligned}$$

Here, the process $(y_t)_{t \in \mathbb{Z}}$ only depends upon the first block of the state $(x_{t,1})_{t \in \mathbb{Z}}$ whose dynamics depend only upon the eigenvalues of the matrix A_{11} . $(x_{t,2})_{t \in \mathbb{Z}}$ is not relevant because of the reduced column rank of C and because of the lower block-triangular structure of the A matrix which implies that there are no feedbacks from $(x_{t,2})_{t \in \mathbb{Z}}$ to $(x_{t,1})_{t \in \mathbb{Z}}$. Consequently, in the above example the second block of the state process $(x_{t,2})_{t \in \mathbb{Z}}$ could have any dynamic property, i.e. could be even an explosive process, since it is not ‘loaded’ into the output. Thus, in case of non-minimal systems, there is typically no one-to-one correspondence between the unit root structures of the output and the state. If the subsystem (A_{11}, B_1, C_1) is minimal then all coordinates of $(x_{t,1})_{t \in \mathbb{Z}}$ ‘appear’ in the output and the unit root structures of $(y_t)_{t \in \mathbb{Z}}$ and $(x_{t,1})_{t \in \mathbb{Z}}$. A precise discussion of this issue is given in Bauer and Wagner (2005, Theorem 3).

Let (A, B, C) be a minimal state space realization of a rational transfer function $k(z)$ and let $(a(z), b(z))$ be a left coprime VARMA realization of the same transfer function, i.e. $k(z) = a^{-1}(z)b(z)$ with $(a(z), b(z))$ left coprime. It holds analogously to the result for stable systems given in Hannan and Deistler (1988, Theorem 1.2.2), that the nonzero eigenvalues of A equal the inverses of the roots of $\det a(z)$. Thus, for minimal state space systems the stability condition $|\lambda_{\max}(A)| < 1$ corresponds to the stability assumption $\det a(z) \neq 0$ for $|z| \leq 1$. The relationship also implies that eigenvalues of A with unit modulus correspond to roots of $\det a(z)$ on the unit circle, i.e. to the unit roots. Considering the inverse transfer function $k^{-1}(z) = b^{-1}(z)a(z) = I_s - zC(I_n - z(A - BC))^{-1}B$ it follows analogously that the condition $\det b(z) \neq 0$ for $|z| < 1$ is equivalent to the condition $|\lambda_{\max}(A - BC)| \leq 1$. The condition $\det b(z) \neq 0$ for $|z| < 1$ is known as miniphase assumption (see Hannan and Deistler 1988, p. 25) and has been imposed in Definition 1 (where it is assumed that $c(z)$ corresponds to the Wold representation). Denote with M_n the set of all rational transfer functions such that $k(0) = I_s$, $\det k(z) \neq 0$ for $|z| < 1$ and that have no pole for $|z| < 1$. Then we can now formally define a *canonical form*, used to achieve identification of state space realizations for all transfer functions in $\bigcup_{n \leq 0} M_n$, as a mapping $\varphi : \bigcup_{n \geq 0} M_n \rightarrow \bigcup_{n \geq 0} S_n$ that attaches a unique state space system $(A, B, C) \in S_n$ to $k(z) \in M_n$.

As already mentioned, the eigenvalues of the matrix A are crucial for the dynamic properties of the state process $(x_t)_{t \in \mathbb{Z}}$ and hence (in a minimal representation also) of the output process $(y_t)_{t \in \mathbb{Z}}$. Let us illustrate the effect of the *eigenvalue structure* for a simple bivariate example, i.e. $y_t \in \mathbb{R}^2$, with a 2-dimensional state process $x_t \in \mathbb{R}^2$ with the only eigenvalues of the matrix A equal to 1. We compare two cases: In the first case $\lambda = 1$ is a simple eigenvalue (i.e. there are two Jordan blocks of size 1 in the Jordan normal form of A , which is hence equal to I_2) and in the second example the eigenvalue $\lambda = 1$ leads to a Jordan block of size 2 in the Jordan normal form of A .⁷ Thus, consider

$$\begin{aligned} y_t &= \begin{bmatrix} C_1 & C_2 \end{bmatrix} x_t + \varepsilon_t \\ \begin{bmatrix} x_{t+1,1} \\ x_{t+1,2} \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,2} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \varepsilon_t, \end{aligned}$$

with $C_1, C_2 \in \mathbb{R}^2$ and $B'_1, B'_2 \in \mathbb{R}^2$. It immediately follows that if the matrix $B = [B'_1, B'_2]'$ has full (row) rank, that the process $(x_t)_{t \in \mathbb{Z}}$ is an I(1) process that is not cointegrated. This is trivially seen by defining $v_t = B\varepsilon_t$ and noting that the covariance matrix of $(v_t)_{t \in \mathbb{Z}}$, $B\Sigma B'$, has full rank when B has full rank, since by assumption $\Sigma > 0$. If both $C_1 \neq 0$ and $C_2 \neq 0$, then both inputs of the state are ‘loaded’ into the output $(y_t)_{t \in \mathbb{Z}}$ and the state space model is minimal.

⁷A full discussion of the eigenvalue structure, its implications and the related so-called *state space unit root structure* is given in Bauer and Wagner (2005, Sections 3-5).

Note that if B does not have full rank, then $B_2 = \tau B_1$ and consequently $v_{t,2} = \tau v_{t,1}$, using $v_t = [v_{t,1}, v_{t,2}]'$. This in turn (using $x_1 = 0$ for simplicity) implies that also $x_{t,2} = \tau x_{t,1}$ and hence $y_t = C_1 x_{t,1} + C_2 x_{t,2} + \varepsilon_t = (C_1 + \tau C_2) x_{t,1} + \varepsilon_t$. Therefore, in case of reduced rank of B , the state space system is not minimal, with a minimal representation given by $y_t = \tilde{C}_1 x_{t,1} + \varepsilon_t$, $x_{t+1,1} = x_{t,1} + B_1 \varepsilon_t$, with $\tilde{C}_1 = C_1 + \tau C_2$ and $\tilde{C}_1 \neq 0$. Thus, (and this observation holds generally) in a minimal representation different Jordan blocks of size 1 correspond to I(1) state processes that are not cointegrated and minimality thus implies full (row) rank of (certain blocks rows of) the B matrix and also places as we have seen some restrictions on (certain block columns of) the C matrix.

Consider now the other example, where the state equation is given by

$$\begin{bmatrix} x_{t+1,1} \\ x_{t+1,2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,2} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \varepsilon_t,$$

Clearly, as before $(x_{t,2})_{t \in \mathbb{Z}}$ is an I(1) process (if $B_2 \neq 0$). For the first coordinate of the state process observe that $\Delta_0(L)(x_{t+1,1}) = x_{t,2} + B_1 \varepsilon_t$, $t \in \mathbb{Z}$ and $\Delta_0(L)^2(x_{t+1,1}) = B_1 \varepsilon_t + (B_2 - B_1) \varepsilon_{t-1}$, $t \in \mathbb{Z}$. Thus, $(x_{t,1})_{t \in \mathbb{Z}}$ is an I(2) process, since the spectral density of $\Delta_0(L)^2(x_{t+1,1})$ at frequency 0 is non-zero for $B_2 \neq 0$. Note that it holds irrespective of B_1 , and also for $B_1 = 0$ in which case $\Delta_0(L)x_{t+1,1} = x_{t,2}$, that $(x_{t,1})_{t \in \mathbb{Z}}$ is an I(2) process for $B_2 \neq 0$. Thus, in the I(2) version of the example minimality only places restrictions on B_2 , i.e. on the block of B corresponding to the I(1) component. Now consider the output equation $y_t = C_1 x_{t,1} + C_2 x_{t,2} + \varepsilon_t$. $(y_t)_{t \in \mathbb{Z}}$ is an I(2) process if $C_1 \neq 0$, irrespective of whether $C_2 \neq 0$ or $C_2 = 0$. If $C_2 = 0$, the process $(x_{t,2})_{t \in \mathbb{Z}}$ nevertheless cannot be dropped from the state space representation since it is input in $(x_{t,1})_{t \in \mathbb{Z}}$.⁸ Altogether we see that minimality puts constraints on (sub-blocks) of the B and C matrices as well as the A matrix.

Summarizing we can draw two conclusions (that hold generally) for minimal representations: First, Jordan blocks of size m correspond to I(m) state processes (with the generalization from 2 to $m > 2$ obvious), where the coordinates of the associated state form a ‘chain’ of increasing integration orders. Second, state coordinates corresponding to different Jordan blocks are not cointegrated. These observations also generalize to unit roots $z = e^{\pm i\omega}$ for $0 < \omega < \pi$, as will be illustrated for MFI(1) processes below.

Let us now turn back to the state space or system equations (3) and let us consider their solutions. For a given value x_1 , the equations can be solved for $t \in \mathbb{Z}$ to obtain (assuming here for brevity for the moment that A is invertible, with the general case being discussed in Bauer and Wagner

⁸The difference to the above example discussing minimality is that here the A matrix is not lower block diagonal.

(2005, Lemma 1))

$$y_t = \begin{cases} CA^{t-1}x_1 + \varepsilon_t + \sum_{j=1}^{t-1} CA^{j-1}B\varepsilon_{t-j}, & t \geq 1, \\ CA^{t-1}x_1 + \varepsilon_t - \sum_{j=t}^0 CA^{j-1}B\varepsilon_{t-j}, & t < 1. \end{cases} \quad (4)$$

This shows that any choice concerning x_1 leads to choosing a unique particular solution of the system equations (3). It is due to the fact that the state equation is an autoregression of order 1 that the set of all solutions to the system equations, for a given process $(\varepsilon_t)_{t \in \mathbb{Z}}$, is so easily described.⁹ Necessarily the solution (4) also shows that the eigenvalues of the matrix A characterize the dynamic behavior (i.e. stationarity or unit root behavior) of $(x_t)_{t \in \mathbb{Z}}$ and hence (in minimal representations) of $(y_t)_{t \in \mathbb{Z}}$.

The observation that the eigenvalues characterize the dynamic behavior allows to make beneficial use of the above mentioned non-uniqueness up to non-singular transformations for minimal realizations by choosing realizations with the A matrix in Jordan normal form (to be precise the part of the A matrix corresponding to the unit modulus eigenvalues) and thus with a particularly simple dynamic structure of the state process. This clearly is possible since the Jordan normal form of the matrix A is similar to the ‘original’ matrix A , i.e. there exists a matrix $T \in \mathbb{R}^{n \times n}$, such that $\text{diag}(J, A_\bullet) = TAT^{-1}$ is such that $|\lambda_{\max}(A_\bullet)| < 1$ and J is in Jordan normal form and corresponds to all unit modulus eigenvalues. To be precise a specifically reordered Jordan normal form will be employed. In the transformed system $(\text{diag}(J, A_\bullet), TB, CT^{-1})$ the properties of the correspondingly transformed state process $(Tx_t)_{t \in \mathbb{Z}}$ are particularly simple. The block-diagonal structure of J decomposes the state vector into sub-vectors corresponding to only one real eigenvalue respectively a pair of complex conjugate eigenvalues and the also decoupled stationary state components corresponding to A_\bullet . This, implies that these sub-vectors of the state are either integrated at only unit root frequency or are stationary. The block-diagonal structure of the transformed A matrix furthermore implies that the subsystems corresponding to the different unit modulus eigenvalues respectively pairs of eigenvalues can be analyzed separately and for the same reason also the stable subsystem can be analyzed separately.¹⁰ Restricting the A matrix to the mentioned form does not lead to a unique representation. Thus, further restrictions have to be imposed on the matrices B and C . These restrictions are formulated for rational unit root processes with arbitrary unit root structures in Bauer and Wagner (2005, Theorems 1 and 2). Rather than developing the necessary identifying restrictions for the general case we will consider the *canonical* representation here only for the three cases of major interest for applied econometric cointegration analysis. These are the

⁹It also follows immediately that considering only zero mean processes $(y_t)_{t \in \mathbb{Z}}$ necessitates $CA^j\mathbb{E}(x_1) = 0$ for $j \in \mathbb{Z}$.

¹⁰This follows since for $A = \text{diag}(A_1, \dots, A_m)$, $B = [B'_1, \dots, B'_m]'$ and $C = [C_1, \dots, C_m]$ partitioned accordingly it follows that the power series coefficients of $k(z) = \Pi(A, B, C)$ fulfill $CA^jB = \sum_{h=1}^m C_h A_h^j B_h$ for $j \geq 0$.

I(1), the MFI(1) and the I(2) case. Discussing these three cases conveys all major ideas already and facilitates the understanding of the general case as discussed in Bauer and Wagner (2005).

The I(1) Case

In the I(1) case the A matrix in the canonical representation is of the following form $A = \text{diag}(I_d, A_\bullet)$ for some $d > 0$ and where $|\lambda_{\max}(A_\bullet)| < 1$, i.e. A_\bullet has only stable eigenvalues.¹¹ Partition $B = [B'_1, B'_\bullet]'$ and $C = [C_1, C_\bullet]$ accordingly with $B_1 \in \mathbb{R}^{d \times s}$ and $C_1 \in \mathbb{R}^{s \times d}$. The subsystem $(A_\bullet, B_\bullet, C_\bullet)$ has a unique stationary solution for $t \in \mathbb{Z}$, which is obtained by setting $x_{1,\bullet} = \sum_{j=1}^{\infty} A_\bullet^{j-1} B_\bullet \varepsilon_{1-j}$ and $x_{1,1}$ is taken to have zero mean but is otherwise unrestricted. The solution to the state space equations in the considered format is then given (as a special case of (4)) by

$$y_t = \begin{cases} C_1 x_{1,1} + C_1 B_1 \sum_{j=1}^{t-1} \varepsilon_{t-j} + w_t, & t \geq 1, \\ C_1 x_{1,1} - C_1 B_1 \sum_{j=t}^0 \varepsilon_{t-j} + w_t, & t < 1, \end{cases} \quad (5)$$

with $w_t = k_\bullet(L) \varepsilon_t = \varepsilon_t + \sum_{j=1}^{\infty} C_\bullet A_\bullet^{j-1} B_\bullet \varepsilon_{t-j}$. Minimality implies (remember the above example) that $C_1 \in \mathbb{R}^{s \times d}$ has full column rank and that $B_1 \in \mathbb{R}^{d \times s}$ has full row rank, which implies $d \leq s$, i.e. the number of I(1) common trends (defined, considering for brevity only $t \in \mathbb{N}$, as $B_1 \sum_{j=1}^{t-1} \varepsilon_{t-j}$) is smaller or equal to the number of variables. The solution (5) for $t \in \mathbb{N}$ extends the Granger representation for I(1) processes as given for AR processes in Johansen (1995, Theorem 4.2). Denote with $C_{1,\perp} \in \mathbb{R}^{s \times (s-d)}$ a full rank matrix such that $C'_{1,\perp} C_{1,\perp} = I_{s-d}$ and $C'_1 C_{1,\perp} = 0$. Then $(C'_{1,\perp} y_t)_{t \in \mathbb{Z}} = (C'_{1,\perp} w_t)_{t \in \mathbb{Z}}$ and is hence stationary, i.e. the column space of $C_{1,\perp}$ spans the cointegrating space. This holds for any zero mean $x_{1,1}$, which reflects the results of Bauer and Wagner (2005, Lemma 1) that starting from an appropriate state space system the property whether the corresponding solution process is a unit root process depends upon the value $x_{1,\bullet}$ only and not upon zero mean $x_{1,1}$.¹²

The I(1) part of the solution in (5) depends only upon the product $C_1 B_1$. In order to define a unique representation of the subsystem corresponding to the I(1) components, i.e. for the system (I_d, B_1, C_1) , a unique decomposition of the product $C_1 B_1$ has to be performed.¹³ One unique decomposition of the product $C_1 B_1$ is described in Bauer and Wagner (2005, Lemma 2), which

¹¹A discussion of the algebraic properties of cointegration in the I(1) case is provided by Neusser (2000).

¹²In order to result in a unit root process, $x_{1,\bullet}$ has to be chosen to render $(w_t)_{t \in \mathbb{Z}}$ stationary. This is only achieved by taking $x_{1,\bullet}$ from the unique stationary solution to the stable subsystem.

¹³Clearly, without further restrictions, the system (I_d, B_1, C_1) is not identified, since for any unitary $T \in \mathbb{R}^{d \times d}$, i.e. $T'T = TT' = I_d$, it holds that $\Pi(I_d, TB_1, C_1 T^{-1}) = \Pi(I_d, B_1, C_1)$, i.e. all these realizations correspond to the same transfer function and describe the same dynamic behavior (which is obvious from (5)). Remember that identification means that if there are two state space realizations (A_1, B_1, C_1) and (A_2, B_2, C_2) of the same transfer function that fulfill all identifying constraints, it holds that $A_1 = A_2$, $B_1 = B_2$, $C_1 = C_2$. Equivalently this can be stated as that starting from an identified realization (A, B, C) of $k(z)$ that fulfills all constraints, the set of all transformation matrices $T \in \mathbb{R}^{d \times d}$ such that also (TAT^{-1}, TB, CT^{-1}) fulfills all constraints contains only $T = I_n$.

requires C_1 to fulfill $C_1' C_1 = I_d$ and B_1 to be *positive upper triangular* and of full row rank. A matrix $M = [m_{i,j}]_{i=1,\dots,c,j=1,\dots,m} \in \mathbb{C}^{c \times m}$ is positive upper triangular (p.u.t.) if there exist indices $1 \leq j_1 < j_2 < \dots < j_c \leq m$, such that $m_{i,j} = 0, j < j_i, m_{i,j_i} \in \mathbb{R}, m_{i,j_i} > 0$, i.e if M is of the form

$$\begin{bmatrix} 0 & \cdots & 0 & m_{1,j_1} & * & \cdots & * \\ 0 & & \cdots & & 0 & m_{2,j_2} & * \\ 0 & & & \cdots & & 0 & m_{c,j_c} & * \end{bmatrix}, \quad (6)$$

where the symbol $*$ indicates unrestricted entries. Note that in case that a real valued matrix is decomposed (as with $C_1 B_1$ in the I(1) case), both matrix factors are real valued. The decomposition, however, also applies to complex valued matrices (see Bauer and Wagner 2005, Lemma 2).¹⁴

Thus, in the I(1) case a unique realization of the subsystem corresponding to the unit roots is of the form: The corresponding A matrix is equal to I_d , the B_1 matrix is p.u.t. with full row rank and the C_1 matrix is normalized to $C_1' C_1 = I_d$. This normalization facilitates the computation of the cointegrating space, which is given by the ortho-complement of the span of C_1 . A unique representation of the total system is then obtained by restricting the stable subsystem $(A_\bullet, B_\bullet, C_\bullet)$, corresponding to the stationary process $(w_t)_{t \in \mathbb{Z}}$ in (5), to be in a canonical form as well. The literature provides numerous possibilities in this respect, e.g. the *balanced canonical form* (see Ober 1996) or the *echelon canonical form* (see Hannan and Deistler 1988, Section 2.5). We consider the echelon canonical form and combining the stable subsystem in echelon canonical form with the unit root subsystem in the unique format described above leads to a unique representation of I(1) state space systems that is well suited for cointegration analysis, since it e.g. immediately leads to a Granger type representation. Also, if one is interested in impulse response analysis and on e.g. wants to place long-run restrictions on the impulse responses, these can be placed on $C_1 B_1$ in a rather straightforward manner.

The MFI(1) Case

The canonical representation for the MFI(1) case is a generalization of the I(1) case to the case of multiple unit roots where some of the unit roots are potentially conjugate complex (i.e. correspond to pairs of complex conjugate eigenvalues of unit modulus of the A matrix). We consider the system in block-diagonal format with the first l blocks corresponding to the unit roots ordered according to increasing unit root frequencies $0 \leq \omega_k < \omega_{k+1} \leq \pi, k = 1, \dots, l-1$ in a format related to the *real Jordan normal form* and the final block being the stationary subsystem. The latter is, as

¹⁴An alternative factorization is given by leaving B_1 unrestricted and by requiring C_1 to fulfill $C_1' C_1 = I_d$ and C_1 *positive lower triangular*, i.e. it is a matrix whose transpose is p.u.t..

before, considered in echelon canonical form:¹⁵

$$y_t = [C_1 \ \cdots \ C_l \ C_\bullet] x_t + \varepsilon_t, \quad (7)$$

$$\begin{bmatrix} x_{t+1,1} \\ \vdots \\ x_{t+1,l} \\ x_{t+1,\bullet} \end{bmatrix} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & A_l & 0 \\ 0 & \cdots & 0 & A_\bullet \end{bmatrix} \begin{bmatrix} x_{t,1} \\ \vdots \\ x_{t,l} \\ x_{t,\bullet} \end{bmatrix} + \begin{bmatrix} B_1 \\ \vdots \\ B_l \\ B_\bullet \end{bmatrix} \varepsilon_t. \quad (8)$$

The unique form of the sub-blocks corresponding to the unit roots depends upon whether the corresponding unit root frequency $\omega_k \in \{0, \pi\}$, i.e. corresponds to a real unit root (± 1), or $0 < \omega_k < \pi$, i.e. corresponds to a pair of complex conjugate unit roots. In case of real unit roots the constraints on the system matrices (A_k, B_k, C_k) are as discussed above, with $A_k = \pm I_{d^k}$ and B_k and C_k fulfilling the constraints formulated above in the I(1) case.

Things are slightly more involved in case of a pair of complex conjugate unit roots $z_k = e^{i\omega_k}$, $\bar{z}_k = e^{-i\omega_k}$, $0 < \omega_k < \pi$, where we have to consider one additional step in the development of a canonical form. From the fact that (A, B, C) as well as $(y_t)_{t \in \mathbb{Z}}$ are real valued it follows after a transformation to the Jordan normal form that not only the blocks in the transformed A matrix that correspond to a pair of conjugate complex unit modulus eigenvalues, z_k and \bar{z}_k , are conjugate complex, but also the correspondingly transformed blocks of B and C are conjugate complex. Consequently the subsystem corresponding to a pair of conjugate complex unit roots is in *complex valued format* given by

$$A_{k,\mathbb{C}} = \begin{bmatrix} \bar{z}_k I_{d^k} & 0 \\ 0 & z_k I_{d^k} \end{bmatrix}, \quad B_{k,\mathbb{C}} = \begin{bmatrix} B_{k,-} \\ \bar{B}_{k,-} \end{bmatrix}, \quad C_{k,\mathbb{C}} = [C_{k,-} \ \bar{C}_{k,-}], \quad (9)$$

with $B_{k,-} \in \mathbb{C}^{d^k \times s}$, $C_{k,-} \in \mathbb{C}^{s \times d^k}$ and where \bar{a} denotes as already used above the complex conjugate of a . A unique realization of the complex subsystem $(\bar{z}_k I_{d^k}, B_{k,-}, C_{k,-})$ is, analogously to the I(1) case discussed above, obtained by positing the constraints $C'_{k,-} C_{k,-} = I_{d^k}$ and $B_{k,-}$ is p.u.t and has full row rank. This, of course, also implies a unique realization of $(z_k I_{d^k}, \bar{B}_{k,-}, \bar{C}_{k,-})$. Based on this unique complex representation, a real valued canonical representation is obtained by transforming the pairs of conjugate complex subsystems $(\bar{z}_k I_{d^k}, B_{k,-}, C_{k,-})$, $(z_k I_{d^k}, \bar{B}_{k,-}, \bar{C}_{k,-})$ to real valued systems using the transformation matrix

$$T_{\mathbb{R},d} := \begin{bmatrix} 1 & & & 1 \\ i & & & -i \\ & 1 & & 1 \\ & i & & -i \\ & & \ddots & \ddots \\ & & & \ddots & \ddots \end{bmatrix} \in \mathbb{C}^{2d \times 2d}. \quad (10)$$

¹⁵The result is again a special case of the canonical representation developed in Bauer and Wagner (2005, Theorem 2).

This results in real valued subsystems (A_k, B_k, C_k) with $A_k = I_{d^k} \otimes Q(\omega_k)$, with $Q(\omega_k) := \begin{bmatrix} \cos(\omega_k) & -\sin(\omega_k) \\ \sin(\omega_k) & \cos(\omega_k) \end{bmatrix}$ and \otimes denoting the Kronecker product. For the other system matrices this transformation leads to

$$B_k = \begin{bmatrix} 2\mathcal{R}(B_{k,-,1}) \\ -2\mathcal{I}(B_{k,-,1}) \\ \vdots \\ 2\mathcal{R}(B_{k,-,d^k}) \\ -2\mathcal{I}(B_{k,-,d^k}) \end{bmatrix}, \quad C_k = [\mathcal{R}(C_{k,-,1}), \mathcal{I}(C_{k,-,1}), \dots, \mathcal{R}(C_{k,-,d^k}), \mathcal{I}(C_{k,-,d^k})],$$

where $B_{k,-,m}$ denotes the m -th row of $B_{k,-}$, $C_{k,-,m}$ denotes the m -th column of $C_{k,-}$ for $m = 1, \dots, d^k$ and \mathcal{R} and \mathcal{I} denote the real respectively imaginary part of a complex quantity.

Consider now only the block corresponding to one unit root with unit root frequency $0 < \omega < \pi$, where we omit here for brevity the subscript k and also set $d = 1$. Then it holds that

$$\begin{bmatrix} x_{t+1,1} \\ x_{t+1,2} \end{bmatrix} = \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix} \begin{bmatrix} x_{t,1} \\ x_{t,2} \end{bmatrix} + B\varepsilon_t,$$

which shows that x_t has unit root structure $(\omega, 1)$ – i.e. is a stochastic cycle to frequency ω – since it holds that $\Delta_\omega(L)(x_t)_{t \in \mathbb{Z}}$ is a stationary process with non-zero spectrum at frequency ω .

Let us next consider (for brevity only for $t \in \mathbb{N}$ and $x_1 = [x'_{1,u}, x'_{1,\bullet}]'$) the solution to the state space system in canonical form, where we consider here the case that $\omega_1 = 0$ and $\omega_l = \pi$:

$$\begin{aligned} y_t = & C_1 B_1 \sum_{j=1}^{t-1} \varepsilon_{t-j} + \sum_{j=1}^{t-1} C_2 (I_{d^2} \otimes Q(\omega_2)^{j-1}) B_2 \varepsilon_{t-j} + \dots + \sum_{j=1}^{t-1} C_{l-1} (I_{d^{l-1}} \otimes Q(\omega_{l-1})^{j-1}) B_{l-1} \varepsilon_{t-j} + \\ & C_l B_l \sum_{j=1}^{t-1} (-1)^{j-1} \varepsilon_{t-j} + \tilde{S}_t + w_t, \end{aligned} \quad (11)$$

with $\tilde{S}_t = C_1 x_{1,1} + C_2 (I_{d^2} \otimes Q(\omega_2)^{t-1}) x_{1,2} + \dots + C_l (-1)^{t-1} x_{1,l}$ summarizing the effect of the state components $x_{1,u} = [x'_{1,1}, \dots, x'_{1,l}]'$ and $w_t = \varepsilon_t + \sum_{j=1}^{\infty} C_\bullet A_\bullet^{j-1} B_\bullet \varepsilon_{t-j}$, using again $x_{1,\bullet} = \sum_{j=1}^{\infty} A_\bullet^{j-1} B_\bullet \varepsilon_{1-j}$.

Thus, the canonical state space representation immediately leads (as before in the I(1) case) to a Granger type representation decomposing the process $(y_t)_{t \in \mathbb{Z}}$ into the stochastic trends $(B_1 \sum_{j=1}^{t-1} \varepsilon_{t-j})$ and cycles $(\sum_{j=1}^{t-1} (I_{d^k} \otimes Q(\omega_k)^{j-1}) B_k \varepsilon_{t-j})$, for $k = 2, \dots, l-1$, $B_l \sum_{j=1}^{t-1} (-1)^{j-1} \varepsilon_{t-j})$ at the different unit root frequencies, the effects of the state $x_{1,u}$ (\tilde{S}_t) and a stationary process (w_t) . The relationship between the state space system matrices and the Granger representation (11) is much simpler than e.g. the relation between AR coefficients and the corresponding Granger representation as derived for the AR case in Johansen and Schaumburg (1999). This implies that it is easier to define a parameterization in the state space framework than in the AR or ARMA framework, see also the next section or Bauer and Wagner (2006) for details.

Bauer and Wagner (2005, Theorem 3) shows that also in the MFI(1), like in the I(1), case the stochastic trends respectively cycles $(x_{t,k})_{t \in \mathbb{Z}}$ are not cointegrated. Thus, cointegration occurs for vectors $\beta \neq 0$ such that $\beta' C_k = 0$ for at least one $k = 1, \dots, l$. Clearly, a vector β can be orthogonal to several matrices C_{k_1}, \dots, C_{k_m} , in which case it annihilates the stochastic trends respectively cycles at several frequencies $\omega_{k_1}, \dots, \omega_{k_l}$.

As already mentioned in the MFI(1) case additionally also so-called *dynamic cointegrating* relationships, i.e. polynomial cointegrating relationships of polynomial degree 1 may occur for unit root frequencies $0 < \omega_k < \pi$, i.e. for the stochastic cycles. Consider $\beta(L) = \beta_0 + \beta_1 L$, with $\beta_0, \beta_1 \in \mathbb{R}^s$ and $\beta_0, \beta_1 \neq 0$. Considering one unit root frequency we obtain

$$(\beta'_0 + \beta'_1 L) C_k x_{t,k} = \beta'_0 C_k \varepsilon_{t-1} + [\beta'_0 C_k (I_{d^k} \otimes Q(\omega_k)) + \beta'_1 C_k] x_{t-1,k}$$

using $x_{t,k} = (I_{d^k} \otimes Q(\omega_k)) x_{t-1,k} + B_k \varepsilon_{t-1}$. Therefore, the polynomial $\beta(L)$ eliminates the stochastic cycle $(x_{t,k})_{t \in \mathbb{Z}}$ in $\beta(L)'(y_t)_{t \in \mathbb{Z}}$ and is a PCIV, since it automatically fulfills condition (ii) of Definition 4, if and only if

$$[\beta'_0 \ \beta'_1] \begin{bmatrix} C_k (I_{d^k} \otimes Q(\omega_k)) \\ C_k \end{bmatrix} = 0.$$

The above matrix of dimension $2s \times 2d^k$, with $d^k \leq s$ (cf. Bauer and Wagner 2005, Lemma 1), has full column rank $2d^k$ due to minimality. This leaves space for a $2(s - d^k)$ -dimensional space of dynamic cointegrating relationships. Similarly as before a vector polynomial $\beta(L)$ can be cointegrating at several unit root frequencies.

The I(2) Case

In the I(2) case the block of the A matrix corresponding to the eigenvalue $z = 1$ has a Jordan normal form that is not equal to the identity matrix, but contains Jordan blocks of size 2, as has been illustrated in the previous sub-section. In the canonical representation a specifically reordered form of the Jordan normal form is used, see (12) below. This specific reordering is chosen to induce that the state vector corresponding to the unit root is ordered block-wise with increasing integration orders from bottom to top. We use for the example the notation used in Bauer and Wagner (2005) for the general case to facilitate reading of that paper, i.e. we have:

$$y_t = \begin{bmatrix} C_1^{1,E} & C_1^{2,G} & C_1^{2,E} & C_\bullet \end{bmatrix} \begin{bmatrix} x_{t,1}^{1,E} \\ x_{t,1}^{2,G} \\ x_{t,1}^{2,E} \\ x_{t,1} \\ x_{t,\bullet} \end{bmatrix} + \varepsilon_t \quad (12)$$

$$\begin{bmatrix} x_{t+1,1}^{1,E} \\ x_{t+1,1}^{2,G} \\ x_{t+1,1}^{2,E} \\ x_{t+1,1} \\ x_{t+1,\bullet} \end{bmatrix} = \begin{bmatrix} I_{d_1} & I_{d_1} & 0 & 0 \\ 0 & I_{d_1} & 0 & 0 \\ 0 & 0 & I_{d_2-d_1} & 0 \\ 0 & 0 & 0 & A_\bullet \end{bmatrix} \begin{bmatrix} x_{t,1}^{1,E} \\ x_{t,1}^{2,G} \\ x_{t,1}^{2,E} \\ x_{t,1} \\ x_{t,\bullet} \end{bmatrix} + \begin{bmatrix} B_1^1 \\ B_1^{2,1} \\ B_1^{2,2} \\ B_\bullet \end{bmatrix} \varepsilon_t,$$

with $x_{t,1}^{1,E} \in \mathbb{R}^{d_1}$, $x_{t,1}^{2,G} \in \mathbb{R}^{d_1}$, $x_{t,1}^{2,E} \in \mathbb{R}^{d_2-d_1}$, $B_1^1 \in \mathbb{R}^{d_1 \times s}$, $B_1^{2,1} \in \mathbb{R}^{d_1 \times s}$, $B_1^{2,2} \in \mathbb{R}^{(d_2-d_1) \times s}$, $C_1^{1,E} \in \mathbb{R}^{s \times d_1}$, $C_1^{2,G} \in \mathbb{R}^{s \times d_1}$, $C_1^{2,E} \in \mathbb{R}^{s \times (d_2-d_1)}$ and $d_1 \leq d_2$.¹⁶ Clearly, in case $d_1 = d_2$ certain components given above are not present. For the discussion here we assume $d_1 < d_2$, but the discussion also makes clear what happens in case $d_1 = d_2$ (as has been the case in the above I(2) example with $d_1 = d_2 = 1$).

In the canonical representation the stable subsystem $(A_\bullet, B_\bullet, C_\bullet)$ is again assumed to be in echelon canonical form with $x_{1,\bullet} = \sum_{j=1}^{\infty} A_\bullet^{j-1} B_\bullet \varepsilon_{1-j}$ corresponding to the unique stationary solution of the stable subsystem. The identifying constraints formulated in Bauer and Wagner (2005, Theorems 1 and 2) are in the I(2) case: $(C_1^E)' C_1^E = I_{d_2}$, where $C_1^E = [C_1^{1,E}, C_1^{2,E}]$, where the full column rank implies $d_2 \leq s$, $(C_1^{2,G})' C_1^{1,E} = 0$, $B_1^2 = [(B_1^{2,1})', (B_1^{2,2})']'$ has full row rank and $B_1^{2,1}$ and $B_1^{2,2}$ are both p.u.t.. These constraints are, of course, closely related to the constraints formulated for the I(1) case. The differences are that full column rank and normalization (i.e. $(C_1^E)' C_1^E = I_{d_2}$) are only formulated for a part of the C_1 matrix (with the E superscript). Constraints on the B_1 matrix are only formulated for B_1^2 , namely full row rank of B_1^2 and p.u.t. format for both sub-blocks $B_1^{2,1}$ and $B_1^{2,2}$.

We now consider the implications of these constraints for the unit root and cointegration properties of $(x_t)_{t \in \mathbb{Z}}$ and $(y_t)_{t \in \mathbb{Z}}$, thereby generalizing the above bivariate I(2) example. Equation (12) implies that $x_{t+1,1}^2 = x_{t,1}^2 + B_1^2 \varepsilon_t$, $t \in \mathbb{Z}$, with $x_{t,1}^2 = [(x_{t,1}^{2,G})', (x_{t,1}^{2,E})']'$. This immediately implies that $(x_{t,1}^2)_{t \in \mathbb{Z}}$ is an I(1) process. Furthermore, since B_1^2 has full row rank $(x_{t,1}^2)_{t \in \mathbb{Z}}$ is not cointegrated. Thus, similarly to the I(1) case, the I(1) coordinates of the state process are not cointegrated. Equation (12) also implies that $x_{t+1,1}^{1,E} = x_{t,1}^{1,E} + x_{t,1}^{2,G} + B_1^1 \varepsilon_t$, $t \in \mathbb{Z}$, and hence $(x_{t,1}^{1,E})_{t \in \mathbb{Z}}$ is an I(2) process. From Bauer and Wagner (2005, Theorem 3) it follows that $(x_{t,1}^{1,E})_{t \in \mathbb{Z}}$ is not cointegrated either, which can also be seen below in (13), due to full row rank of $B_1^{2,1}$. Thus, the state vector $(x_{t,1})_{t \in \mathbb{Z}}$ is given in a form where its sub-blocks $(x_{t,1}^j)_{t \in \mathbb{Z}}$, for $j = 1, 2$ are integrated of order $3 - j$ and are not cointegrated. Thus, the number of I(1) common trends is d_2 , which like in the I(1) case is smaller or equal than s , and the number of I(2) common trends is given by d_1 , also smaller or equal than s .

Considering again for notational brevity only the solution for $t \in \mathbb{N}$ we obtain for the integrated components of the state vector:

$$x_{t,1}^{1,E} = x_{1,1}^{1,E} + (t-1)x_{1,1}^{2,G} + B_1^{2,1} \sum_{i=1}^{t-1} \sum_{j=1}^{i-1} \varepsilon_{i-j} + B_1^1 \sum_{j=1}^{t-1} \varepsilon_{t-j} \quad (13)$$

¹⁶The reason for as well as the need for different double superscripts for the blocks of C_1 and $x_{t,1}$ on the one hand and for B_1^2 only becomes clear when looking at higher order integrated processes (see Bauer and Wagner 2005, Theorem 2).

$$\begin{aligned}
x_{t,1}^{2,G} &= x_{1,1}^{2,G} + B_1^{2,1} \sum_{j=1}^{t-1} \varepsilon_{t-j} \\
x_{t,1}^{2,E} &= x_{1,1}^{2,E} + B_1^{2,2} \sum_{j=1}^{t-1} \varepsilon_{t-j}
\end{aligned} \tag{14}$$

This implies for the solution y_t , $t \in \mathbb{N}$, using $w_t = \varepsilon_t + \sum_{j=1}^{\infty} C_{\bullet} A_{\bullet}^{j-1} B_{\bullet} \varepsilon_{t-j}$ for the stationary component and $\tilde{S}_t = C_1^{1,E} [x_{1,1}^{1,E} + (t-1)x_{1,1}^{2,G}] + C_1^{2,G} x_{1,1}^{2,G} + C_1^{2,E} x_{1,1}^{2,E}$ for the effect of the state $x_{1,1}$:

$$y_t = C_1^{1,E} B_1^{2,1} \sum_{i=1}^{t-1} \sum_{j=1}^{i-1} \varepsilon_{i-j} + \underbrace{(C_1^{1,E} B_1^1 + C_1^{2,G} B_1^{2,1} + C_1^{2,E} B_1^{2,2})}_{C_1 B_1} \sum_{j=1}^{t-1} \varepsilon_{t-j} + \tilde{S}_t + w_t \tag{15}$$

From the solution (15) the cointegration properties of $(y_t)_{t \in \mathbb{Z}}$ can be immediately deduced. First, since both $C_1^{1,E}$ and $B_1^{2,1}$ are full rank matrices, a vector $\beta \neq 0, \beta \in \mathbb{R}^s$ reduces the integration order from 2 to (at least) 1 if and only if $\beta' C_1^{1,E} = 0$. Reduction of the integration order from 2 to 0 necessitates some additional orthogonality constraints to be fulfilled, namely $\beta' (C_1^{2,G} B_1^{2,1} + C_1^{2,E} B_1^{2,2}) = 0$. The dimension of this space depends upon the process considered, but can be easily determined for any given process by simple orthogonality constraints. E.g. in case that $C_1 B_1 \in \mathbb{R}^{s \times s}$ has full rank, no cointegration that reduces the integration order from 2 to 0 exists. The above representation (15) that follows straightforwardly from the state space equations immediately displays cointegration and is analogous to the representation developed for the autoregressive case in Johansen (1997, (10), p. 437).

It remains to discuss polynomial cointegration. Bauer and Wagner (2005, Theorem 4) states that a necessary condition for a polynomial $\beta(L) = \beta_0 + \beta_1 L$ to be a polynomial cointegrating vector is

$$[\beta'_0, \beta'_1] \begin{bmatrix} C_1 J_1 \\ C_1 \end{bmatrix} = 0. \tag{16}$$

In the I(2) case this leads to the orthogonality constraint

$$[\beta'_0, \beta'_1] \begin{bmatrix} C_1^{1,E} & C_1^{1,E} + C_1^{2,G} & C_1^{2,E} \\ C_1^{1,E} & C_1^{2,G} & C_1^{2,E} \end{bmatrix} = 0 \tag{17}$$

Since minimality implies that the above matrix has full column rank the solution space to the above equation is of dimension $2(s - d_1 - d_2)$. Now, the above orthogonality constraints are only necessary conditions, since they do not incorporate the item (ii) of Definition 4, which excludes polynomials that are (in the I(2) case) multiples of the differencing filter $\Delta_0(L) = 1 - L$. In the I(2) case a polynomial fulfills item (ii) if and only if it holds that

$$[\beta'_0, \beta'_1] \begin{bmatrix} I_s \\ I_s \end{bmatrix} \neq 0,$$

i.e. if it does not fulfill a certain orthogonality constraint. Thus, the intersection of the orthogonal complement of the left kernel of the above matrix with the solution space of (17) characterizes all polynomial cointegrating relationships. Similar ideas hold also for the case of unit root processes with arbitrary unit root structure, as is discussed in detail in Bauer and Wagner (2008). I.e. in other words, the characterization problem of polynomial cointegration is transformed into a static orthogonality problem. That paper also discusses the notion of *minimum degree* polynomial cointegrating vectors. This is necessary since for a PCIV $\beta(L)$ any vector polynomial $p(L)\beta(L)$, where $p(L)$ is a scalar polynomial, is also a PCIV, as long as the pre-multiplication with $p(L)$ does not invalidate condition (ii) of Definition 4. In the I(2), respectively I(m), case this just requires that $p(1) \neq 0$. Clearly, only the set of minimum degree PCIVs of a certain order is relevant. The characterization of minimum degree PCIVs can also be formulated as an orthogonality problem.

3.2 Statistical Theory

Compared to structure theory, which is developed for unit root processes with arbitrary unit structures, statistical theory is in a relatively nascent state with results available only for some cases. Pseudo maximum likelihood estimation theory is developed to some extent for the MFI(1) case and for the I(1) case a computationally simple so-called subspace algorithm for the estimation of the system matrices as well as for order estimation (n) and testing for the number of common trends (d) has been developed. Here we only very briefly discuss the available results and refer the reader to the original papers for details. For pseudo ML estimation see Bauer and Wagner (2006) and for subspace algorithm cointegration analysis see Bauer and Wagner (2002, 2009).

Pseudo ML Estimation for MFI(1) Processes

The major purpose of the developed canonical form, probably more important than allowing for an easy understanding of the unit root, cointegration and polynomial cointegration properties of $(y_t)_{t \in \mathbb{Z}}$, is that it allows to define a *parameterization*. Formally, in our context a (finite-dimensional) parameterization of M_n is a bijective mapping from $\psi : T \rightarrow M_n$, with $T \subset \mathbb{R}^m, m < \infty$ such that $\tau \in T \rightarrow \psi(\tau) = k(z, \tau) = \Pi(A(\tau), B(\tau), C(\tau)) \in M_n$, i.e. a parameterization assigns in a bijective fashion a transfer function to a parameter vector. A canonical form is an important ingredient for a parameterization since it attaches a unique state space realization to any $k(z) \in M_n$. As we have seen, the canonical representation places restrictions on the system matrices, i.e. not all entries in (A, B, C) are free parameters. The free parameters in the canonical representation are collected in the parameter vector $\tau \in \mathbb{R}^m$. For a detailed discussion see Hannan and Deistler (1988).

In our context, for given unit root frequencies and unit root structure (to be precise, for given state space unit root structure as defined in Bauer and Wagner 2005), which in the MFI(1) case means the multiplicities d_1, \dots, d_l of the unit roots, there are no free parameters in the part of the A matrix corresponding to the unit roots and the corresponding blocks C_k and B_k , $k = 1, \dots, l$, have to fulfill certain constraints (ortho-normality, orthogonality, p.u.t.). Similarly, using the echelon canonical form for the stable subsystem places restrictions on $(A_\bullet, B_\bullet, C_\bullet)$.

It is known since Hazewinkel and Kalman (1976) that the set of transfer functions M_n cannot be parameterized continuously. Clearly, however, continuity and even differentiability of a parameterization are desirable properties. This, since continuity implies that when the estimated parameters are ‘close’ to the true parameters also the corresponding transfer functions are ‘close’ in a certain sense and hence the estimated model exhibits characteristics that resemble the characteristics of the true but unknown model. Differentiability allows to use gradient based methods for optimization of the likelihood and also allows to use (in case that the true parameter vector is an interior point of the parameter set) expansions around the true parameter vector to derive asymptotic distributions. Thus, in order to have *piece-wise* continuous parameterizations for subsets of M_n , this set is partitioned, i.e. $M_n = \bigcup_\theta M_n^\theta$, according to some index θ that summarizes the partitioning, as discussed next for the MFI(1) case, with a continuous and differentiable parameterizations of M_n^θ .

To be concrete let us discuss this for the considered MFI(1) case: Let n be fixed and consider also a fixed unit root structure $\Omega = ((\omega_1, 1), \dots, (\omega_l, l))$. Furthermore, let the multiplicities of the unit roots be denoted as $d = (d_1, \dots, d_l)$. Furthermore, the structure of the p.u.t. indices (which describe the position of the first non-zero entry in each row) for the matrices $B_{k,-}$ is collected in θ_{put} . Finally, with θ_\bullet we denote the structure (multi-)index characterizing the echelon canonical representation of $(A_\bullet, B_\bullet, C_\bullet)$, see Hannan and Deistler (1988, Theorem 2.5.2, p. 61). Combine these in $\theta = (n, \Omega, d, \theta_{put}, \theta_\bullet)$. Then, the set M_n is partitioned according to the defined parameter θ as just defined. Given the focus on the cointegration properties it appears natural to partition M_n according to the unit root properties as summarized by Ω and d as well as according to the partitioning induced by the echelon canonical form for the stationary subsystem.¹⁷

The parameter vector is partitioned in three components, i.e. $\tau = (\tau_C, \tau_B, \tau_\bullet)$ and of course there are also $s \times (s + 1)/2$ parameters to be estimated in Σ . The precise dimensions of the parameter vectors are discussed in Bauer and Wagner (2006). Statistical analysis is based on $(-2/T)$ the

¹⁷More details and properties of a parameterization are contained in Bauer and Wagner (2006), which uses the factorization referred to in footnote (14). The underlying topological properties of the parameterization discussed there are a special case of results contained in an older version of Bauer and Wagner (2005), which is available from the author upon request.

Gaussian pseudo (log-)likelihood,

$$L(\tau, \Sigma | y_1, \dots, y_T) = \log \det(\Sigma) + \frac{1}{T} \hat{\varepsilon}_t(\tau) \Sigma^{-1} \hat{\varepsilon}_t(\tau), \quad (18)$$

with $\hat{\varepsilon}_t(\tau)$ defined from

$$\hat{\varepsilon}_t(\tau) = y_t - C(\tau)x_t(\tau) \quad (19)$$

$$x_{t+1}(\tau) = A(\tau)x_t(\tau) + B(\tau)\hat{\varepsilon}_t(\tau). \quad (20)$$

The pseudo ML estimators $\hat{\tau}$ and $\hat{\Sigma}$ are defined as the minimizers of the function (18) over some compact $\Lambda \subset T$.

Under a set of appropriate assumptions, which include correctly specified structure indices n, d and the θ_{put} , as well as assumptions on $(\varepsilon_t)_{t \in \mathbb{Z}}$ (to be a strictly stationary MDS with constant conditional variance and finite fourth moments) and some technical assumptions on $k(z)$ Bauer and Wagner (2006, Theorem 1) establish consistency of $\hat{\tau}$.

Furthermore, if the true parameter point, τ_0 say, is an interior point of the parameter space also the asymptotic distribution is available (see Bauer and Wagner 2006, Theorem 2). As is probably not a surprise, the parameters in C are estimated super-consistently, i.e. $T(\hat{\tau}_\gamma - \tau_{\gamma,0}) \rightarrow B(\Omega, d)$. Here $B(\Omega, d)$ is used to denote a ‘complicated’ vector of Brownian motions whose precise form depends upon the unit root structure Ω and the multiplicities d . The other parameters, τ_B and τ_\bullet , are estimated at the standard rate \sqrt{T} and are asymptotically normally distributed. All parameters converge jointly. Availability of the asymptotic distribution allows to construct hypothesis tests on the parameters. An important ingredient missing to date is a consistent order estimation criterion, i.e. a consistent estimator of the order n as well as tests for the number of common trends respectively cycles (i.e. for d).

Subspace Algorithm Cointegration Analysis for I(1) Processes

For the special case of I(1) processes ‘complete’ statistical analysis including order estimation as well as testing for the number of common trends has been developed in Bauer and Wagner (2002, 2009) using an alternative estimation approach referred to in the literature as subspace algorithms.

Subspace algorithms originated in the engineering literature of the 1980s (see e.g. Larimore 1983, Van Overschee and DeMoor 1994, Verhaegen 1994) and consequently have been mainly used in a stationary context. The exception is Aoki (1987, Chapter 9), whose procedure however lacks a thorough statistical foundation and can be shown to be inefficient for stationary processes. Bauer and Wagner (2002) extend and modify the so-called CCA (Canonical Correlation Analysis) algorithm of Larimore (1983) from the stationary case to the I(1) case.

Despite the fact that subspace algorithm cointegration analysis is the to date only published strand, and hence most easily available part, of the state space model cointegration analysis research agenda we briefly present the main idea, since subspace algorithms may not be too well-known in the econometrics literature. As has been seen above, pseudo ML estimation proceeds by obtaining estimates of the system matrices $(\hat{A}, \hat{B}, \hat{C})$, from which in turn a corresponding estimate of the state \hat{x}_t can be constructed. The idea of subspace algorithms is to reconsider the system equations assuming that the in fact unobserved state process were observed or a consistent estimate is available, \hat{x}_t say. In that case the matrices in the system equations $y_t = C\hat{x}_t + \varepsilon_t$, $\hat{x}_{t+1} = A\hat{x}_t + B\varepsilon_t$ can be estimated by least squares methods. E.g. the output equation $y_t = C\hat{x}_t + \varepsilon_t$ can be used to estimate \hat{C} and $\hat{\varepsilon}_t$, which then in turn can be used to estimate \hat{A} and \hat{B} from $\hat{x}_{t+1} = A\hat{x}_t + B\hat{\varepsilon}_t + \xi_t$, with ξ_t denoting here the regression residual. Thus, in order to transform the estimation problem to such a simple regression problem a consistent estimator of the state has to be constructed and the asymptotic effect of using \hat{x}_t in place of x_t has to be studied.

There are several ways of doing that and CCA type algorithms base the estimation of the state on the canonical correlations between $Y_{t,f}^+ = [y'_t, y'_{t+1}, \dots, y'_{t+f-1}]'$ on $Y_{t,p}^- = [y'_{t-1}, y'_{t-2}, \dots, y'_{t-p}]'$ for appropriately chosen indices $f, p \geq n$ (with in general $p \rightarrow \infty$ for $T \rightarrow \infty$ at a suitable rate; hence in this step an autoregressive approximation is performed, compare also Remark 1). Denote $S_{++} = \frac{1}{T_{f,p}} \sum_{t=p+1}^{T-f} Y_{t,f}^+ (Y_{t,f}^+)', S_{--} = \frac{1}{T_{f,p}} \sum_{t=p+1}^{T-f} Y_{t,p}^- (Y_{t,p}^-)'$ and $S_{+-} = \frac{1}{T_{f,p}} \sum_{t=p+1}^{T-f} Y_{t,f}^+ (Y_{t,p}^-)'$. The algorithm is based on the canonical correlations between $Y_{t,f}^+$ on $Y_{t,p}^-$, in our notation given by $S_{++}^{-1/2} S_{+-} S_{--}^{-1/2}$.¹⁸ It holds that for the canonical correlation matrix the number of non-zero eigenvalues (i.e. canonical correlations) is asymptotically equal to the system order n . This fact is exploited for consistent order estimation. Furthermore, the number of eigenvalues asymptotically equal to 1 equals the number of common trends. For a chosen order \hat{n} an estimate of the state is (essentially) computed from the singular value decomposition of the canonical correlation matrix considering only the first n canonical correlations. I.e. consider a singular value decomposition of $S_{++}^{-1/2} S_{+-} S_{--}^{-1/2} = \hat{U} \hat{\Sigma} \hat{V}' = \hat{U}_n \hat{\Sigma}_n \hat{V}_n' + \hat{R}_n$, with $\hat{\Sigma}_n \in \mathbb{R}^{n \times n}$ and the other matrices of appropriate dimensions and where R_n captures the contribution of all neglected canonical correlations. Then, an estimate of the state is given by $\hat{x}_t = \hat{\Sigma}_n \hat{V}_n' S_{--}^{-1/2} Y_{t,p}^-$.¹⁹ Based on the estimated state, as discussed above, the system matrices can be estimated. Tests for the number of common trends can

¹⁸Note also that the coefficient matrix of a regression of $Y_{t,f}^+$ on $Y_{t,p}^-$ is given by $\hat{\beta}_{f,p} = S_{+-} S_{--}^{-1}$, which of course is related to the canonical correlation matrix by pre- and post-multiplication with $S_{++}^{-1/2}$ and $S_{--}^{1/2}$. Different subspace algorithms differ, amongst other things, by their choice of weighting matrices with which the OLS coefficient matrix $\hat{\beta}_{f,p}$ is pre- and post-multiplied. Note that a specific choice of weighting matrices also leads to a specific 'solution' of the identification problem.

¹⁹Bauer and Wagner (2002) discuss a modified estimate of the state taking into account the number of common trends, d in our notation. We abstain from detailing this modification here.

be based on the eigenvalues of \hat{A} , since asymptotically the d largest eigenvalues are equal to 1 (cf. Bauer and Wagner 2005, Theorem 5). Tests can also be based on a reduced rank regression. Thus, subspace algorithms are a computationally extremely cheap (one SVD and two OLS regressions) alternative to pseudo maximum likelihood estimation for cointegrated I(1) VARMA processes in the state space framework.

Both simulation evidence as well as empirical applications have shown that the outlined procedure performs well. Wagner (2004) uses subspace algorithm cointegration analysis, and for comparison the methods of Bierens (1997) and Johansen (1995), to test for and to estimate the cointegrating space for the one-sector neoclassical growth model (see also King, Plosser and Rebelo 1988, King *et al.* 1991). The estimated cointegrating space is closer (in the sense of the gap distance) to the cointegrating space implied by economic theory than the estimates obtained with the other methods.

Bauer and Wagner (2009) use the method to investigate the *expectations hypothesis of the term structure* from a cointegration perspective, which has some tradition by now in the cointegration literature, see e.g. Campbell and Shiller (1987) or Hall *et al.* (1992). The expectations hypothesis, dating back to Fisher (1896) and Hicks (1946)²⁰, has – when cast into a set-up amenable to cointegration analysis – clear implications for cointegration: The interest rates to all different maturities are considered to be I(1), whereas all *spreads*, i.e. interest rate differentials, are stationary. I.e. for s interest rates of different maturities the dimension of the cointegrating space implied by theory is $s - 1$, with a basis given by the vectors of the form $[1, 0, \dots, 0, -1, 0, \dots, 0]'$ with the -1 entry running from the second to the last position. Note that in terms of state space models as considered, this simply implies that the ortho-complement of the cointegrating space, i.e. $C_1 \in \mathbb{R}^{s \times 1}$ is proportional to $[1, \dots, 1]'$, which in other words just states the fact that the underlying stochastic trend is loaded in all interest rates with the same coefficient.

Using US and German interest rate (with maturities from 1 to 12 months) and bond returns data (with maturities from 1 to 6 years) Bauer and Wagner (2009) find strong support for the outlined expectations hypothesis in that indeed they find $s - 1$ cointegrating relationships, or equivalently only one stochastic trend, for the US and German interest rate data and the US bond data. Applying the Johansen (1995) VAR method leads to the conclusion of a much lower dimensional cointegrating space. Also, an estimate of the underlying factor (using here the word factor in the sense of underlying theoretical finance models like Cox *et al.* 1985, Vasicek 1977) is immediately available as $\hat{x}_{t,1}$.

The two mentioned applications show that using state space models, which allows to model

²⁰For a ‘modern’ exposition see Shiller (1990).

VARMA processes, may make a difference in cointegration analysis. Remember that for both applications there are theoretical reasons why considering VARMA processes might be relevant, in the growth example this is the observation that the log-linearized solutions of economic models often follow VARMA processes, and in the term structure example VARMA processes may be relevant due to temporal aggregation of the underlying (instantaneous) interest rate modelled as a diffusion process in the mentioned finance models. Clearly, in both cases the argument for VARMA processes stems from theoretical models that need not be supported by the data. Nevertheless, it may be useful to take findings from theoretical considerations into account when setting out for empirical econometric analysis.

4 Open Issues, Summary and Conclusions

From the above discussed results the open issues can be seen immediately, by considering the set of results one would ultimately want to have available when performing cointegration analysis. With respect to structure theory the major open issue is the inclusion of exogenous variables (which includes, of course, deterministic variables), or in other words the consideration of unit root VARMAX instead of unit root VARMA processes. This leads to considering state space models of the form

$$y_t = Cx_t + Mz_t + \varepsilon_t \quad (21)$$

$$x_{t+1} = Ax_t + Nz_t + B\varepsilon_t, \quad (22)$$

with $(z_t)_{t \in \mathbb{Z}}$ containing the exogenous variables. Depending upon the nature of $(z_t)_{t \in \mathbb{Z}}$ this raises many interesting questions, both structure theoretical and statistical. Many economic variables contain in addition to the stochastic trend also deterministic trend components, thus at least allowing for certain deterministic trend components is relevant.²¹

With respect to statistical theory it is clear that many things are missing, since not even for the three discussed cases (I(1), MFI(1) and I(2)) complete statistical analysis ranging from order estimation to parameter estimation to testing for the number of common trends and cycles is available. Some results might be relatively easy to achieve (e.g. the distribution of the impulse response coefficients, given the distribution of the parameters in (A, B, C)) whereas others might be more difficult (like showing consistency and deriving the asymptotic distribution of estimators for processes with general unit root structures). It is a major purpose of the developed canonical form to set the stage for parameter estimation for processes with general unit root structures.

²¹Note that a non zero mean of the processes is easily captured by non zero $x_{1,1}$

Alternatively, or as a complement, due to their computational simplicity one may also want to extend subspace algorithms beyond the $I(1)$ case.

Altogether this shows that cointegration analysis with state space models is a field with many interesting questions still to be formulated let alone solved and we hope that this summary paper spurs the interest of some researchers to work in this interesting field.

Acknowledgements

I would like to thank the Deutsche Statistische Gesellschaft, in particular Karl Mosler and Wilfried Seidel, for the invitation to give the Gumbel lecture at the *Statistische Woche* in Wuppertal in October 2009. Furthermore, cointegration analysis with state space models is a joint research agenda with Dietmar Bauer, with whom to collaborate has been a pleasure for more than a decade now. I thank Robert Kunst and Klaus Neusser for helpful comments. Finally, I would like to thank Manfred Deistler, who is ‘responsible’ for my turning to econometrics.

References

- Ahn, S. K. and Reinsel, G. C.: 1994, Estimation of partially nonstationary vector autoregressive models with seasonal behavior, *Journal of Econometrics* **62**, 317–350.
- Aoki, M.: 1987, *State Space Modeling of Time Series*, Springer, New York.
- Aoki, M. and Havenner, A.: 1989, A method for approximate representation of vector valued time series and its relation to two alternatives, *Journal of Econometrics* **42**, 181–199.
- Aoki, M. and Havenner, A.: 1991, State space modeling of multiple time series, *Econometric Reviews* **10**(1), 1–59.
- Bauer, D. and Wagner, M.: 2002, Estimating cointegrated systems using subspace algorithms, *Journal of Econometrics* **111**, 47–84.
- Bauer, D. and Wagner, M.: 2005, A canonical form for unit root processes. *Mimeo*.
- Bauer, D. and Wagner, M.: 2006, Asymptotic theory of pseudo maximum likelihood estimators for multiple frequency $I(1)$ processes. *Mimeo*.
- Bauer, D. and Wagner, M.: 2007, Autoregressive approximations of multiple frequency $I(1)$ processes. *Mimeo*.
- Bauer, D. and Wagner, M.: 2008, Polynomial cointegration reconsidered. *Mimeo*.

- Bauer, D. and Wagner, M.: 2009, Using subspace algorithm cointegration analysis: simulation performance and application to the term structure, *Computational Statistics and Data Analysis* **53**, 1954–1973.
- Bierens, H.: 1997, Nonparametric cointegration analysis, *Journal of Econometrics* **77**, 379–404.
- Box, G. E. P. and Tiao, G. C.: 1977, A canonical analysis of multiple time series, *Biometrika* **64**, 355–365.
- Campbell, J. Y.: 1994, Inspecting the mechanism: An analytical approach to the stochastic growth model, *Journal of Monetary Economics* **33**, 463–506.
- Campbell, J. Y. and Shiller, R. J.: 1987, Cointegration and tests of present value models, *Journal of Political Economy* **95**, 1062–1088.
- Cox, J.C., Ingersoll, J.E., Ross, S.A.: 1985, A theory of the term structure, *Econometrica* **53**, 385–407.
- Dickey, D. A. and Fuller, W. A.: 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* **49**, 1057–1072.
- Durbin, J. and Koopman, S.J.: 2001, *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.
- Engle, R. F. and Granger, C. W. J.: 1987, Cointegration and error correction: representation, estimation and testing, *Econometrica* **55**, 251–276.
- Fisher, I.: 1896, Appreciation and interest, *Publications of the American Economic Association*, 23–29 and 88–92.
- Granger, C. W. J.: 1981, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* **16**, 121–130.
- Granger, C. W. J. and Lee, T.: 1989a, Investigation of production, sales and inventory relationship using multicointegration and non-symmetric error correction models, *Journal of Applied Econometrics* **4**, 145–159.
- Granger, C. W. J. and Lee, T.: 1989b, Multicointegration, in G. Rhodes and T. Fomby (eds), *Advances in Econometrics: Cointegration, Spurious Regression and Unit Roots*, JAI Press, Greenwich, 71–84.

- Granger, C. W. J. and Newbold, P.: 1974, Spurious regressions in econometrics, *Journal of Econometrics* **2**, 111–120.
- Gregoir, S.: 1999, Multivariate time series with various hidden unit roots, part I, *Econometric Theory* **15**, 435–468.
- Gregoir, S. and Laroque, G.: 1994, Polynomial cointegration: estimation and test, *Journal of Econometrics* **63**, 183–214.
- Hall, A. D., Anderson, H. M. and Granger, C. W. J., 1992, A cointegration analysis of treasury bill yields, *Review of Economics and Statistics* **74**, 116–126.
- Hannan, E. J. and Deistler, M.: 1988, *The Statistical Theory of Linear Systems*, John Wiley, New York.
- Harvey, A.: 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Hazewinkel, M. and Kalman, R. E.: 1976, Invariants, canonical forms and moduli for linear, constant, finite dimensional, dynamical systems, in G. Marchesini and S. K. Bitter (eds), *Lecture Notes in Economics and Mathematical Systems*, Springer, New York, 48–60.
- Hicks, J.R., 1946: *Value and Capital*, 2nd Ed, Oxford University Press, Oxford.
- Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B.: 1990, Seasonal integration and cointegration, *Journal of Econometrics* **44**, 215–238.
- Johansen, S.: 1995, *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford University Press, Oxford.
- Johansen, S.: 1997, Likelihood analysis of the I(2) model, *Scandinavian Journal of Statistics* **24**, 433–462.
- Johansen, S. and Schaumburg, E.: 1999, Likelihood analysis of seasonal cointegration, *Journal of Econometrics* **88**, 301–339.
- Kailath, T.: 1980, *Linear Systems*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Kim, C.-J. and Nelson, C. R.: 1999, *State Space Models with Regime Switching*, MIT Press, Cambridge (Mass.).

- King, R. G., Plosser, C. I. and Rebelo, S.: 1988, Production, growth and business cycles II: new directions, *Journal of Monetary Economics* **21**, 309–341.
- King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W.: 1991, Stochastic trends and economic fluctuations, *American Economic Review* **81**, 819–840.
- Krämer, W.: 1981, Ordinary least squares estimation of simultaneous equation systems with trended data, presented at the Econometric Society European Meeting, Amsterdam.
- Larimore, W.E., 1983: System identification, reduced order filters and modelling via canonical variate analysis, in H. S. Rao and P. Dorato (eds), *Proceedings 1983 American Control Conference 2*, 445–451.
- Lee, H.: 1992, Maximum likelihood inference on cointegration and seasonal cointegration, *Journal of Econometrics* **54**, 1–47.
- Lewis, R. and Reinsel, G. C.: 1985, Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**, 393–411.
- Lütkepohl, H. and Claessen, H.: 1997, Analysis of cointegrated VARMA processes, *Journal of Econometrics* **80**, 223–229.
- Nelson, C. R. and Plosser, R. I.: 1982, Trends and random walks in macroeconomic time series, *Journal of Monetary Economics* **10**, 139–162.
- Neusser, K.: 2000, An algebraic interpretation of cointegration, *Economics Letters* **67**, 273–281.
- Ober, R.: 1996, Balanced canonical forms, in S. Bittanti and G. Picci (eds), *Identification, Adaptation, Learning: The Science of Learning Models from Data*, Springer, Berlin, 120–183.
- Park, J. Y. and Phillips, P. C. B.: 1988, Statistical inference in regressions with integrated regressors: part I. *Econometric Theory* **4**, 468–497.
- Park, J. Y. and Phillips, P. C. B.: 1989, Statistical inference in regressions with integrated regressors: part II. *Econometric Theory* **5**, 95–131.
- Phillips, P. C. B.: 1986, Understanding spurious regressions in econometrics, *Journal of Econometrics* **33**, 311–340.
- Phillips, P. C. B. and Hansen, B. C.: 1990, Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies* **57**, 99–125.

- Saikkonen, P.: 1991, Asymptotically efficient estimation of cointegrating regressions, *Econometric Theory* **7**, 1–21.
- Saikkonen, P.: 1992, Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**, 1–27.
- Saikkonen, P. and Luukkonen, R.: 1997, Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics* **81**, 93–126.
- Shiller, R.J.: 1990, The term structure of interest rates, in B. Friedman and F. H. Hahn (eds), *Handbook of Monetary Economics*, Elsevier, Amsterdam, 627–722.
- Sims, C. A., Stock, J. H. and Watson, M. W.: 1990, Inference in linear time series models with some unit roots, *Econometrica* **58**, 113–144.
- Stock, J. H. and Watson, M. W.: 1988, Testing for common trends, *Journal of the American Statistical Association* **83**, 1097–1107.
- Stock, J. H. and Watson, M. W.: 1993, A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**, 783–820.
- Van Overschee, P. and DeMoor, B.: 1994, N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems, *Automatica* **30**, 75–93.
- Vasicek, O.A.: 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**, 177–188.
- Verhaegen, M.: 1994, Identification of the deterministic part of mimo state space models given in innovations form from input-output data, *Automatica* **30**, 61–74.
- Wagner, M.: 2004, A comparison of Johansen’s, Bierens’ and the subspace algorithm method for cointegration analysis, *Oxford Bulletin of Economics and Statistics* **66**, 399–424.
- Yap, S. F. and Reinsel, G. C.: 1995, Estimation and testing for unit roots in a partially non-stationary vector autoregressive moving average model, *Journal of the American Statistical Association* **90**, 253–267.
- Yoo, B.: 1986, Multi-cointegrated time series and generalized error correction models. *Technical report*, Department of Economics, University of California San Diego.
- Yule, G. U.: 1926, Why do we sometimes get nonsense-correlations between time series? - A study in sampling and the nature of time series, *Journal of the Royal Statistical Society* **89**, 1–63.

Zellner, A. and Palm, F. C.: 1974, Time series analysis and simultaneous equation econometric models, *Journal of Econometrics* **2**, 17–54.

Author: Martin Wagner

Title: Cointegration Analysis with State Space Models

Reihe Ökonomie / Economics Series 248

Editor: Robert M. Kunst (Econometrics)

Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

© 2010 by the Department of Economics and Finance, Institute for Advanced Studies (IHS),
Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • <http://www.ihs.ac.at>
